Sparsity-promoting migration accelerated by message passing

Xiang LI and Felix J.Herrmann* EOS-UBC*

SUMMARY

Seismic imaging via linearized inversion requires multiple iterations to minimize the least-squares misfit as a function of the medium perturbation. Unfortunately, the cost for these iterations are prohibitive because each iteration requires many wave-equation simulations, which without direct solvers require an expensive separate solve for each source. To overcome this problem, we use dimensionality-reduction to decrease the size of seismic imaging problem by turning the large number of sequential shots into a much small number of simultaneous shots. In our approach, we take advantage of sparsifying transforms to remove source crosstalk resulting from randomly weighting and stacking sequential shots into a few super shots. We also take advantage of the fact that the convergence of large-scale sparsity-promoting solvers can be improved significantly by borrowing ideas from message passing, which are designed to break correlation built up between the linear system and the model iterate. In this way, we arrive at a formulation where we run the sparsity-promoting solver for a relatively large number of iterations. Aside from leading to a significant speed up, our approach had the advantage of greatly reducing the memory imprint and IO requirements. We demonstrate this feature by solving a sparsity-promoting imaging problem with operators of reverse-time migration, which is computationally infeasible without the dimensionality reduction.

INTRODUCTION

Nowadays, industry tends to favor migration methods which vary from 'time' or 'depth' to 'one-way wave-equation' and to 'reverse time' and so on. All these migration methods are aimed at imaging subsurface reflectors, however, none of them provide correct amplitude information. Least-squares migration has been introduced to overcome this amplitude problem by inverting the system iteratively while minimizing the least-squares as a function of the medium perturbation (see e.g. Rickett, 2003; Plessix and Mulder, 2004). As such, least-squares migration provides us in principle with artifact-free true-amplitude images even in situations where the background velocity model contains minor errors (Kühl and Sacchi, 2003). Because each source requires a separate partial-differential equation (PDE) solve, this leads to wave simulation costs that increase linearly with the number of sources. This requirement proves to be prohibitive in practice where datasets contain hundreds-of-thousands of sources or more and this explains the slow adaptation of this technology.

To reduce these computational cost, we follow earlier ideas on random source encoding (Morton and Ober, 1998; Romero et al., 2000; Herrmann et al., 2009; Neelamani et al., 2010) to turn the large 'overdetermined' system of equations of the discretized linearized Born scattering operator into an 'underdetermined' system by dimensionality reduction. During this dimensionality reduction, we randomly combining all sequential shots into small subsets of simultaneous shots (Li and Herrmann, 2010; Krebs et al., 2009; Haber et al., 2010). We subsequently use ideas from compressive sensing (CS, Candès et al., 2006; Donoho, 2006; Mallat, 2009) to remove the source crosstalk by solving a sparsity-promoting program (SPGℓ_1, Berg and Friedlander, 2008) that exploits curvelet-domain sparsity on the model.

While this approach certainly leads to a significant reduction of the problem size, which leads to a corresponding reduction in the number of PDE solves required by each iteration, the convergence of SPGℓ_1 becomes a limiting factor especially when targeting the removal of small source-interference that lie close to the null-space of the forward modeling operator. To address this challenging issue, we borrow ideas from approximate message passing (AMP Donoho et al., 2009). Instead of adding a message term, which depends on unrealizable assumptions on the forward model, we draw new subsets of supershots after solving each subproblem of SPGℓ_1. Under certain conditions, the selection of new supershots is statistically to approximate message passing (Montanari, 2010). As a consequence, correlations between the model iterate and forward model are removed, which leads to a significant improvement of the convergence.

Our outline is as follows: First, we introduce how to apply dimensionality reduction to seismic imaging. Second, we discuss efficient imaging with sparsity promotion, improvements by message passing, and a practical adaptation of message passing to sparsity-promoting imaging. We conclude by applying the proposed method to a seismic imaging problem with well-log derived velocity model.

THEORY

Dimensionality reduced least squares migration: Least-squares migration involves inversion of the linearized (time-harmonic) acoustic Born-scattering modeling operator and has the following separable form:

$$\underset{\delta\mathbf{m}}{\text{minimize}}\ \frac{1}{2K}\|\delta\mathbf{D}-\nabla\mathscr{F}[\mathbf{m}_0;\mathbf{Q}]\delta\mathbf{m}\|_F^2 = \frac{1}{2K}\sum_{i=1}^{K}\|\delta\mathbf{d}_i-\nabla\mathscr{F}_i[\mathbf{m}_0;\mathbf{q}_i]\delta\mathbf{m}\|_2^2. \quad (1)$$

In this expression, $\nabla\mathscr{F}[\mathbf{m}_0;\mathbf{Q}]$ is the linear Born-scattering modeling operator with background velocity model $\mathbf{m}_0$ for all sources collected in the columns of $\mathbf{Q}$, the matrix $\delta\mathbf{D}\in\mathbb{C}^{N_f N_r N_s}$ contains the observed wavefield with N_f, N_r, and N_s the number of angular frequencies, receiver, and source positions. The vector $\delta\mathbf{m}\in\mathbb{R}^M$ is the unknown model perturbation, with M the number of gridpoints. Since each angular frequency and sequential source can be treated independently, index $i=1\cdots K$ represents each single experiment based on one frequency and one shot.

Least-squares migration is challenging because each iteration to solve Equation 1 typically requires 4K PDE solves: two for the action of $\mathscr{F}_i$ and the other two for its adjoint. Thus, the inversion costs grow linearly with the number of monochromatic source experiments, multiplied by the number of matrix-vector multiplies required by the solver. Consequently, iterative methods require multiple passes thought the whole data to evaluate the action of the Born scattering operator and its adjoint.To reduce costs, we replace the sequential shots with a small number of simultaneous shots ($K' \ll K$) supershots yielding

$$\underset{\delta\mathbf{m}}{\text{minimize}} \frac{1}{2K'} \sum_{i=1}^{K'} \|\delta\mathbf{D}\mathbf{w}_i - \nabla\mathscr{F}_i[\mathbf{m}_0;\mathbf{Q}\mathbf{w}_i]\delta\mathbf{m}\|_2^2 \\ = \frac{1}{2}\|\underline{\delta\mathbf{D}} - \nabla\mathscr{F}_i[\mathbf{m}_0;\underline{\mathbf{Q}}]\delta\mathbf{m}\|_F^2, \qquad (2)$$

where $\{\underline{\delta\mathbf{D}}, \underline{\mathbf{Q}}\} := \{\mathbf{DW}, \mathbf{QW}\}$ (Moghaddam and Herrmann, 2010; Haber et al., 2010; van Leeuwen et al., 2011). We obtain the dimensionality reduction of Equation 1 by multiplying the observed wavefield and sources from the right with a random Gaussian tall matrix $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_{K'}]$. In that way, we can easily see that the number of PDE solvers required for each iteration of the solution of Equation 2 decreases by a factor of K'/K(see also Herrmann et al., 2009, for details). While random-amplitude source encoding and subsampling, allows us to reduce the number of sources, the required source superposition leads to source crosstalk, which we need to remove to enhance the image quality.

Efficient imaging with sparsity promoting: To mitigate source-crosstalk artifacts caused by the randomized subsampling, we take advantage of compressive sensing where sparse N-long vectors $\mathbf{x}$ can be recovered from incomplete measurements $\mathbf{b} = \mathbf{Ax}$, where $\mathbf{b} \in \mathbb{C}^n$ and $\mathbf{A} \in \mathbb{C}^{n\times N}$ is the sensing matrix with $n \ll N$. According to CS (Candès et al., 2006; Donoho, 2006; Mallat, 2009), recovery is possible for certain matrices $\mathbf{A}$. Using this result, we argue that we can solve Equation 2 with the following sparsity-promoting program known as Basis Pursuit (BP):

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2}\|\mathbf{x}\|_1 \quad \text{s.t.} \quad \underline{\delta\mathbf{D}} = \nabla\mathscr{F}[\mathbf{m};\underline{\mathbf{Q}}]\mathbf{S}^H\mathbf{x}. \qquad (3)$$

In this expression, $\mathbf{S}^H$ is the inverse curvelet transform that we use to represent the model perturbation—i.e., $\delta\mathbf{m} = \mathbf{S}^H\mathbf{x}$. Following van den Berg and Friedlander (2008), we solve this sparsity-promotion program with a root-finding method on the Pareto curve. This approach corresponds to solving a series of LASSO (Tibshirani, 1997) subproblems that read

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2}\|\underline{\delta\mathbf{D}} - \nabla\mathscr{F}[\mathbf{m};\underline{\mathbf{Q}}]\mathbf{S}^H\mathbf{x}\|_F^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \le \tau \qquad (4)$$

and where τ is one-norm for constraint. See Figure 1, for a typical solution in relation to the underlying Pareto curve. While this framework leads to an efficient algorithm to solve large-scale sparsity-promoting programs, the convergence of the algorithm stalls, which leads to unsatisfactory results as reported by Herrmann and Li (2011).

Improvement of the convergency by message passing: Each sub-problem solved by SPGℓ_1 involves relatively expensive gradient updates, $\mathbf{x}^{t+1} = \mathbf{x}^t + \mathbf{A}^T(\underline{\delta\mathbf{D}} - \mathbf{Ax}^t)$, with the sensing matrix $\mathbf{A} = \nabla\mathscr{F}[\mathbf{m};\underline{\mathbf{Q}}]\mathbf{S}^H$. These gradient updates are followed by an orthogonal projection of the model iterate to ℓ_1-norm ball of size τ (Berg and Friedlander, 2008). As observed by Herrmann (2012), this procedure leads to a build up of correlations between the sensing matrix and the model iterates. These correlations are the result of misidentified coefficients in the vector for the model iterate, which are difficult to remove. To remove these correlations between the sensing matrix and the model iterates, we follow the spirit of message passing by drawing a new set of simultaneous sources after each LASSO sub-problem of SPGℓ_1 is solved. (See Figure 1. As reported by Montanari (2010), and by Herrmann (2012) selection of a new copy of the sensing matrix and data has the same effect as including a message term. The advantage of this approach is that is does not rely on the assumptions underlying AMP that are very difficult to meet in our application. See Figure 1 and Algorithm 1, which illustrates the principle of our proposed algorithm.

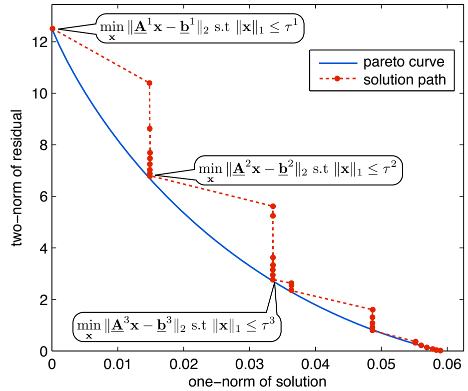


Figure 1: Series of LASSO subproblems with renewals for the collections of supershots (adapted from Berg and Friedlander (2008)).

Result: Estimate for the model $\widetilde{\mathbf{x}}$

$\mathbf{x}_0 \longleftarrow \mathbf{0}$; // initial model

$\mathbf{k} \longleftarrow \mathbf{0}$; // initial counter

$\{\underline{\delta\mathbf{D}}, \underline{\mathbf{Q}}\} := \{\mathbf{DW}, \mathbf{QW}\}$; // random shots

while $\|\mathbf{x}_0 - \widetilde{\mathbf{x}}\|_2 \ge \varepsilon$ **do**

  $k \longleftarrow k+1$; // increase counter

  $\widetilde{\mathbf{x}} \longleftarrow \mathbf{x}_0$; // update warm start

  $\mathbf{x}_0 \longleftarrow \text{Solve}(\text{minimize}_{\mathbf{x}} \frac{1}{2}\|\underline{\delta\mathbf{D}} - \nabla\mathscr{F}[\mathbf{m};\underline{\mathbf{Q}}]\mathbf{S}^H\mathbf{x}\|_F^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \le \tau)$; // solve the subproblem

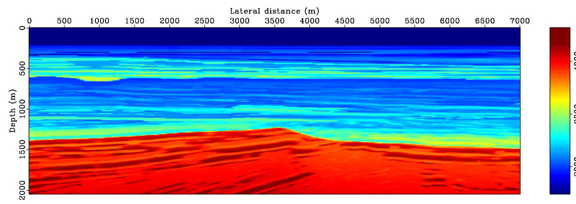
  $\{\underline{\delta\mathbf{D}}, \underline{\mathbf{Q}}\} := \{\mathbf{DW}, \mathbf{QW}\}$; // do redraws

end

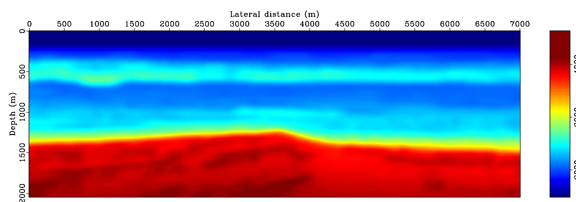
Algorithm 1: sparsity promoting recovery with approximated message passing

EXAMPLE

To test the performance of our algorithm in a realistic setting, we use BG compass model (Figure 2(a)) as the true reference model, which is constrained by well-log information. This sedimentary model has two big unconformities and a sharp velocity 'kick-back' under the first unconformity, which is challenging for migration. The initial model (Figure 2(b)) for the migration is generated by applying a low-pass filter on the true model.



(a)



(b)

Figure 2: BG compass model. **(a)** True model. **(b)** Initial model.

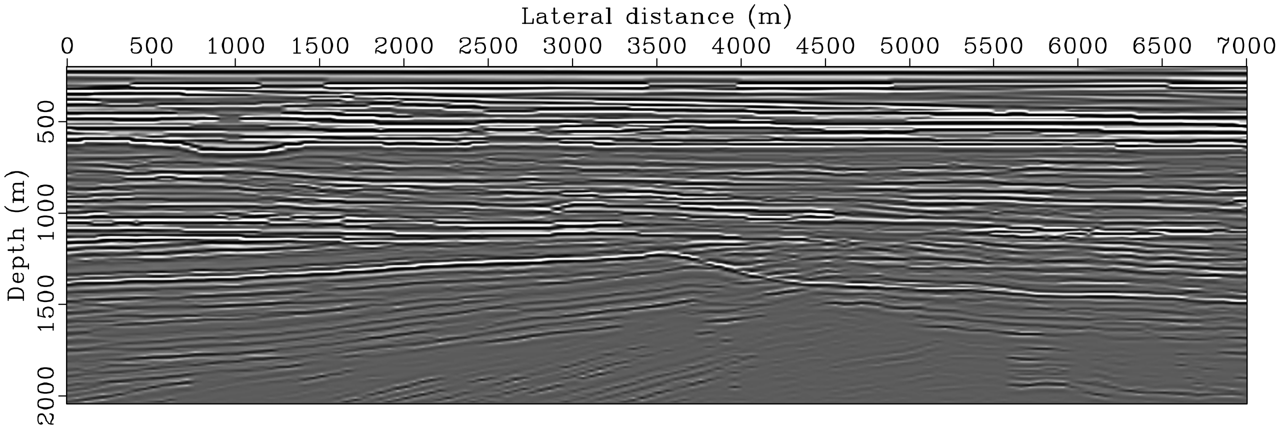
We use 10 random frequencies in our simulations selected from the interval $20-50\,\mathrm{Hz}$ with a source function given by a $30\,\mathrm{Hz}$ Ricker wavelet. We parametrize the velocity perturbation on a 409×1401 grid with a gridsize of $5\,\mathrm{m}$. We use the Helmholtz solver to generate data from 350 source positions sampled at an interval of $20\,\mathrm{m}$ and 701 receivers sampled with an interval of $10\,\mathrm{m}$. Since we want to focus on improvements in convergence, we avoid issues related to the linearization by generating data with the linearized Born scattering operator. We only use three randomly selected simultaneous shots and we solve 50 subproblems both with and without selecting new independent copies of the sensing matrix and data. In either case, the costs are about 500 SPG$\ell1$ iterations. Therefore, the total computational cost of these two experiments is equivalent to solving the imaging problem (cf. Equation 1) for all 350 sequential shots with $3--5$ LSQR iterations. The result of this exercise is summarized in figure 3 and clearly shows significant improvements from including the message term. Not only is the crosstalk removed more efficiently but the reflectors are also better imaged, in particular at the deeper parts of the model where recovery without redraws is not able to image the events.

CONCLUSIONS

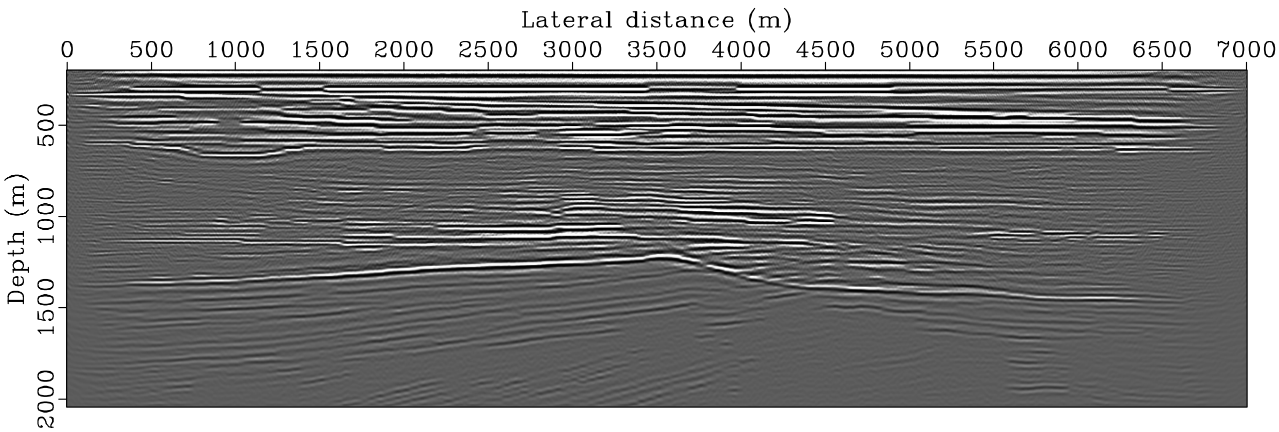
In this abstract, we introduced an efficient algorithm to solve the linearized image problem. The combination of dimensionality-reduction, curvelet-domain sparsity-promotion, and message-passing, via drawing new copies of the sensing matrix and data, leads to a remarkable speedup of convergence and improved image quality. The explanation for these improvements lies mainly in our ability to prevent correlations to build up between the sensing matrix and the model iterates. In this way, we are able to work with very small subsets of simultaneous source experiments while still being able to remove the source interferences. Our results are remarkable because we are able to get very high quality images for least-squares imaging problems that are too expensive to be solved using all data during each iteration. For other applications of message passing, we refer the reader to other contributions by the authors these proceedings.

ACKNOWLEDGMENTS

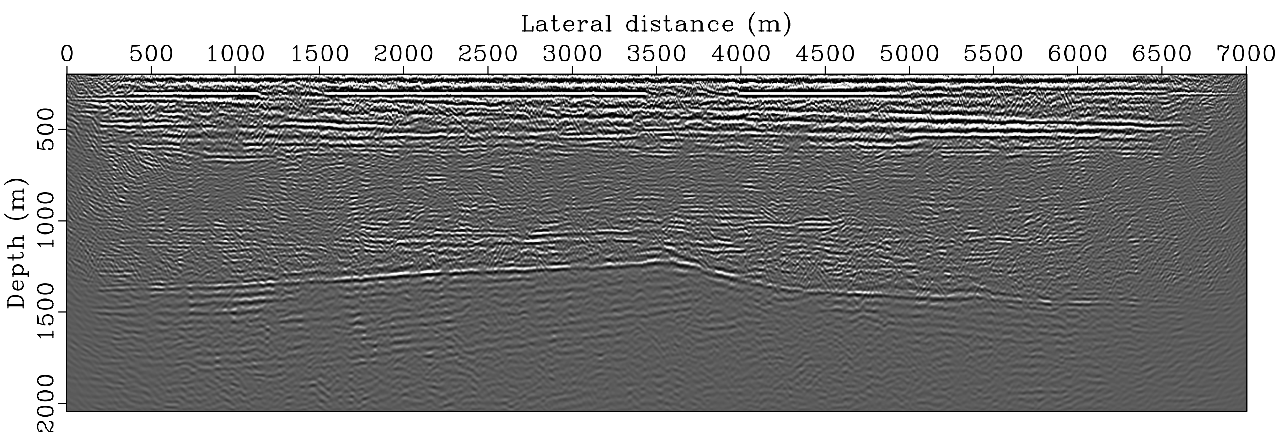
We would like to thank Charles Jones from BG for providing us with the BG Compass model. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, BGP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.



(a)



(b)



(c)

Figure 3: Dimensionality-reduced sparsity-promoting imaging from random subsets of three simultaneous shots and ten frequencies. We used the background velocity-model plotted in Figure 2(b) **(a)** True perturbation given by the difference between the true velocity model and the smoothed initial model plotted in Figure 2(b). **(b)** Imaging result with 'messaging'. **(c)** The same but without 'messaging'.

REFERENCES

Berg, E. v., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Technical Report 2, Department of Computer Science, University of British Columbia.

Candès, E., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Comm. Pure Appl. Math., **59**, 1207–1223.

Donoho, D., A. Maleki, and A. Montanari, 2009, Message passing algorithms for compressed sensing: Proceedings of the National Academy of Sciences, **106**, 18914–18919.

Donoho, D. L., 2006, Compressed sensing: IEEE Trans. Inform. Theory, **52**, 1289–1306.

Haber, E., M. Chung, and F. J. Herrmann, 2010, An effective method for parameter estimation with PDE constraints with multiple right hand sides: Technical Report TR-2010-4, UBC-Earth and Ocean Sciences Department. (revised August 2011).

Herrmann, F. J., 2012, Pass on the message: recent insights in large-scale sparse recovery: Presented at the EAGE technical program, EAGE, EAGE.

Herrmann, F. J., Y. A. Erlangga, and T. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35.

Herrmann, F. J., and X. Li, 2011, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **NA**, NA. (To appear in Geophysical Prospecting).

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Kühl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for avp/ava inversion: Geophysics, **68**, 262273.

Li, X., and F. J. Herrmann, 2010, Full-waveform inversion from compressively recovered model updates: , SEG, 1029–1033.

Mallat, S. G., 2009, A Wavelet Tour of Signal Processing: the Sparse Way: Academic Press.

Moghaddam, P. P., and F. J. Herrmann, 2010, Randomized full-waveform inversion: a dimenstionality-reduction approach: SEG, Expanded Abstracts, **29**, 977–982.

Montanari, A., 2010, Graphical models concepts in compressed sensing: CoRR, **abs/1011.4328**.

Morton, S. A., and C. C. Ober, 1998, Faster shot-record depth migrations using phase encoding: SEG Technical Program Expanded Abstracts, SEG, 1131–1134.

Neelamani, R. N., C. E. Krohn, J. R. Krebs, J. K. Romberg, M. Deffenbaugh, and J. E. Anderson, 2010, Efficient seismic forward modeling using simultaneous random sources and sparsity: Geophysics, **75**, WB15–WB27.

Plessix, R., and W. Mulder, 2004, Frequency-domain finite difference amplitude-preserving migration: Geoph. J. Int., **157**, 975–987.

Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: Geophysics, **68**, no. 4, 1371–1379.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, no. 2, 426–436.

Tibshirani, R., 1997, Regression shrinkage and selection via the LASSO: J. Royal. Statist. Soc B., **58**, 267–288.

van den Berg, E., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.

van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: International Journal of Geophysics, article ID: 689041.